

OM of: US-09-805-550-4 to: GenBank: out\_format: pfs

Date: Aug 31, 2002 7:35 AM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

#### Command line parameters:

```
-MODEL-firmes+pn.model -DEV-xlh  
-Q/cgnt2.1/USPTO.spool/US09805550/runal_29082002.160819.375/app-query.fasta.1.897  
-Db-GenBank -OFMT-fasta -SUFFIX-oligop2n.rge -GAPOP-4.500  
-GAPEXT-0.050 -MINMATCH-0.100 -LOOPE-0.000 -LOOPEXT-0.000  
-GAPOP-4.500 -GAPEXT-0.050 -XGAPOP-60.000 -XGAPEXT-60.000  
-FGAPOP-6.000 -FGAPEXT-7.000 -YGAPOP-60.000 -YGAPEXT-60.000  
-DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-oligo  
-TRANS-human40.cdi -LIST-45 -DOCALLIGN-200 -THR_SCORE-quality  
-THR_MIN-1 -ALIGN-15 -MODE-LOCAL -OUTFMT-pfs -NORM-ext  
-HEMPSIZE-500 -MINLEN-0 -MAXLEN-200000000  
-USER-US09805550.ecgnt2.1.4959 -NCPU-6 -ICPU-3 -LONGLOC  
-DEV-TIMEOUT-120 -WARN-TIMEOUT-30 -NO_XLPHY -WAIT -THREADS-1
```

#### Search information block:

```
Query: US-09-805-550-4  
Query length: 368  
Database: GenBank: *  
Database sequences: 1797656  
Database length: 1873333701  
Search time (sec): 3229.260000
```

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Sequence	Strid	Orig	ZScore	EScore	len	Documentation
gb_pat:AR153820	+	368.00	6342.60	0.0	1702	AR153820 Sequence 3 from patent
gb_pat:AX314196	+	56.00	950.35	8.5e-45	353	AX314196 Sequence 7181 from patent
gb_scs:G71829	+	54.00	920.56	6.5e-43	312	G71829 A9481814FMO17 maize leaf
gb_hlg:AP004214	+	23.00	348.44	4.8e-11	70957	AP004214 Oryza sativa chromos
gb_pl:AV063103	+	21.00	341.09	1.2e-10	1129	AV063103 Arabidopsis thaliana
gb_pl:AV034912	+	21.00	339.14	1.6e-10	1518	AV034912 Arabidopsis thaliana
gb_pl:AC010793	+	21.00	312.29	5.0e-09	90149	AC010793 Genomic sequence for
gb_hlg:AP003881	+	18.00	256.99	4.6e-06	110775	AP003881 Oryza sativa chrom
gb_pl:AV037181	+	15.00	233.96	8.8e-05	1378	AV037181 Arabidopsis thaliana
gb_pl:AV058196	+	15.00	233.75	9.1e-05	1426	AV058196 Arabidopsis thaliana
gb_pl:DCRA231T	+	15.00	235.53	9.3e-05	1466	Y12014 Daucus carota mRNA for
gb_pl:DCRA231T	+	15.00	235.41	9.5e-05	1496	Y12014 Daucus carota mRNA for
gb_pl:FC09	+	15.00	206.86	0.0037	114498	AC006341 Arabidopsis thalian
gb_pl:LES243875	+	14.00	218.10	0.0009	1498	AJ243875 Lycopersicon esculent
gb_pl:AT4D18	+	14.00	191.88	0.0252	80442	AC010924 Arabidopsis thaliana
gb_pat:AR153819	+	13.00	200.65	0.0081	1533	AR153819 Sequence 1 from patent
gb_pl:OS065330	+	12.00	157.02	2.20	83646	AB005248 Arabidopsis thaliana
gb_pl:AR005248	+	12.00	157.02	2.20	83646	AB005248 Arabidopsis thaliana
gb_pl:CMS01ATV	+	11.00	171.01	0.3664	720	AL113387 Botrytis cinerea strai
gb_pl:CMS01D80	+	11.00	166.28	0.6727	1479	AF132147 Drosophila melanogast
gb_in:AR132147	+	11.00	165.49	1.12	1667	AY039562 Arabidopsis thaliana
gb_pl:AV039562	+	11.00	162.32	1.12	2697	AF136606 Drosophila melanogast
gb_in:AR136606	+	11.00	153.02	3.68	11090	AC018037 Drosophila melanogast
gb_hlg:AC018037	+	11.00	138.29	24.36	103904	AC021640 Arabidopsis thaliana
gb_pl:ARAC021640	+	11.00	138.29	24.36	103904	AC021640 Arabidopsis thaliana
gb_in:AC010667	+	11.00	130.88	63.05	159570	AC010667 Drosophila melanogast
gb_in:AR003844	+	10.00	145.91	8.94	411	AJ270958 Lycopersicon esculentus
gb_pl:LES270958	+	9.00	145.91	8.94	411	AJ270958 Lycopersicon esculentus
gb_in:AF250297	+	9.00	145.91	9.17	170	AF250297 Chironomus balatonicus
gb_in:AF250297	+	9.00	145.91	9.17	170	AF250297 Chironomus balatonicus
gb_in:AF250301	+	9.00	145.11	9.21	171	AF250301 Chironomus plumosus hg
gb_pr:HG295373	+	9.00	145.11	9.21	171	AF250301 Chironomus plumosus hg
gb_ba:BSR282971	+	9.00	141.25	16.66	345	282971 Bacterial sp. partial 16
gb_ba:AF360632	+	9.00	140.31	18.80	398	AF360632 Desulfohalobium sp. CMF
gb_ba:US008395	+	9.00	139.04	22.14	483	US008395 Unknown sulfate-reducing
gb_ba:AF121085	+	9.00	138.82	23.76	499	AF121085 Uncultured bacterium C
gb_ba:US008389	+	9.00	138.67	23.22	511	US008389 Unknown sulfate-reducing
gb_ba:US008394	+	9.00	138.67	23.22	511	US008394 Unknown sulfate-reducing

gb_ba:US008385	+	9.00	138.60	23.41	516	US008385 Unknown sulfate-reduc
gb_ba:AF360637	+	9.00	138.27	24.44	543	AF360637 Uncultured bacteriu
gb_ba:AF360642	+	9.00	138.27	24.44	543	AF360642 Uncultured delta pr
gb_ba:AF360638	+	9.00	138.25	24.48	544	AF360638 Uncultured bacteriu
gb_ba:AF360641	+	9.00	138.25	24.48	544	AF360641 Uncultured bacteriu
gb_ba:AF360639	+	9.00	138.24	24.51	545	AF360639 Uncultured bacteriu
gb_ba:AB062690	+	9.00	138.17	24.74	551	AB062690 Uncultured bacteriu

  

seq_name: gb_pat:AR153820	
seq documentation block:	
LOCUS	AR153820 1702 bp DNA linear PAT 08-AUG-2001
DEFINITION	Sequence 3 from patent US 6235972.
ACCESSION	AR153820
VERSION	AR153820.1 GI:15121352
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	1 (bases 1 to 1702)
AUTHORS	Manjani, P.B. and Tagliani, L.
TITLE	Maize Rad23 genes and uses thereof
JOURNAL	Patent: US 6235972-A 3 22-MAY-2001;
FEATURES	Location/Qualifiers
source	1..1702
BASE COUNT	490 a 373 c 432 g 407 t
ORIGIN	

  

alignment_scores:	
Quality:	368.00
Ratio:	1.000
Percent Similarity:	100.000
Percent Identity:	100.000

  

alignment_block:	
US-09-805-550-4 x AR153820 ..	
Align seg 1/1 to: AR153820 from: 1 to: 1702	

  

1	MetlyleuThrValylsThrLeuLysGlyThrHisPheGluLeuArg	17
106	ATGAGCTGAGCGGTGAGACCTCAAGGAGGACCTGAGATCCGGGT	155
17	1GlnProAsnAspThrIleMetAlaValLysLysAsnIleGluIleG	34
156	GCAGCCCAACGACACGATATGCTGTGAAGAGATATGAGAGATAC	205
34	1InclLysAspSerTyrProTrrGlyGlnGlnLeuLeuIlePheAsn	50
206	AAAGGAAAGACACCTATCCATGGGGCCACACCTGATTTTCATGGA	255
51	1LysValLeuLysAspGlnSerThrLeuGlnGlnAsnLysValAsnG	67
256	AAAGCTTGAAGATGAATACATTGGAAGAGATTAAGTCATATGGA	305
67	1GlyPheLeuValAlaMetLeuSerLysGlyLysThrSerGlySerThr	84
306	TGGGTTTCTGATGTCATGCTAGTAAGGTAAACATCGTCAACTG	355
84	1LysThrSerSerGlnHisSerAsnThrProAlaThrArgGlnAlaPro	100
356	GAACCTTCATCTTCCAGCACTCAAACTCTCTCAACAGCAGGACCT	405
101	1ProLeuGlnAlaProGlnGlnAlaProGlnAlaProValAlaProIle	117
406	CTCTTGAAGGCCCCACACAGCTCTCAACCCCGGTGGACCAATTAC	455
117	1TrpSerGlnProGluGlyLeuProAlaGlnAlaProAsnThrHisAsp	134
456	AACTTTCACACCTGGAAGACTTCTGACAGGACCTTAACACATGACA	505
134	1SnaIlaIaSerAsnLeuSerGlyArgAsnValAspThrIleIleAsn	150

|||||  
506 ATGGGGCAATCTTCTGTGGAGGAATGTGACACATAATTAAC 555  
151 GlnleuMetGluMetGlyGlySerTrpAspLysAspLysValGln 167  
556 CAGCTAATGAGATGGGTGGGGCAGTTGGACAAAGATTAAGTCCAAAG 605  
167 GAlaLeuArgAlaAlaTrpAsnAsnProGluArgAlaValGluTrpLeu 184  
606 GGCTCTCCGTCGGCTTACAAACCCGGAAGCTGTGTGAATACCTCT 655  
184 YrSerGlyTrpProValThrAlaGluIleAlaValProIleGlyGln 200  
656 ACTGGATTCAGTACAGCTGAATGCTGTTCATTTGGTGTCAA 705  
201 GAlaAlaAsnThrTrpAspArgAlaProThrGlyGluAlaGlyLeuSerG 217  
706 GGGGCAACACACATGATCGAGCTCTACTGGGAGAGCTGGTCTCTGG 755  
217 YrLeProAsnThrAlaProLeuAspLeuPheProGlnGlyAlaSerAsn 234  
756 GATTCCAACACACCGCTCCACTAGATCTTTCCCGCAGGGGCTTCCAAAT 805  
234 lAgGlyGlyAlaGlyGlyGlyProLeuAspPheLeuArgAsnPro 250  
806 CTGGAGGT 855  
251 GlnPheGlnAlaValArgGluMetValHisThrAsnProGlnIleLeuG 267  
856 CAGTTTCAGACAGTTAGGGAATGTGTCATACAAATCCACAAATTTTGA 905  
267 nPrometLeuValGluLeuSerLysGlnAsnProGlnIleLeuArgLeu 284  
906 GCCATGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 955  
284 lGluGluAsnHisAspGluPheLeuGlnLeuLeuAsnGluProPheGlu 300  
956 TTGAGAGCAATCATGATGATGATTTCTTCAGTTACTAAATGAGCCCTTGA 1005  
301 GlyGlyGluGluAspPheLeuAspGlnProGluGluAspGluMetProH 317  
1006 GCGGAGAGGGGAGATTCTTTAGACCAACTGAGAGGATGAATGCCCA 1055  
317 SalAlaSerValThrProGluGluGlnGluAlaIleGlyArgLeuGln 334  
1056 TGCATTATGTTTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1105  
334 eMetGlyPheAspArgAlaArgValIleGluAlaPheLeuAlaCysAsp 350  
1106 CCATGGGCTTGCACAGACAGCGCTTATTGAAGCATTTTACCTGCGAT 1155  
351 ArgAsnGluGluLeuAlaAlaAsnTrpLeuLeuGlnHisAlaGlyGluG 367  
1156 AGGAACGAGAGAGCTAGACGAACACTATCTCTTGAGCATGCTGTGAGA 1205  
367 uAsp 368  
1206 AGAT 1209

seq\_name: gb\_pat:AX314196

seq\_documentation\_block:

LOCUS AX314196 353 bp DNA linear PAT 14-DEC-2001

DEFINITION Sequence 7181 from Patent WO0190366.

ACCESSION AX314196

VERSION AX314196.1 GI:17898333

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (sites) Mammalia: Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Leach,M.D. and Shimkets,R.A.

TITLE Human polynucleotides and polypeptides encoded thereby  
JOURNAL Patent: WO 0190366-A 7181 29-NOV-2001;  
Curagen Corporation (US)  
FEATURES location/Qualifiers  
source 1.353  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 92 a 71 c 100 g 90 t  
ORIGIN

alignment\_scores:  
Quality: 56.00 Length: 56  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-805-550-4 x AX314196 ..

Align seg 1/1 to: AX314196 from: 1 to: 353

253 GlnAlaValArgGluMetValHisThrAsnProGlnIleLeuGlnProMe 269  
|||||  
85 CAGCAGCTTATGAGGAATGTGTCATACAAATCCACAAATTTTGCAGCTAT 134  
269 lLeuValGluLeuSerLysGlnAsnProGlnIleLeuArgLeuGln 286  
135 GCTCGTGTGAGTTGAGCAGAGCAAGATCTCAATCTTAAGGTGATGAGG 184  
286 lAsnHisAspGluPheLeuGlnLeuLeuAsnGluProPheGluGly 302  
185 AGAATCATGATGATGATTTCTTCAGTTACTAAATGAGCCCTTGAAGCGGA 234  
303 GlnGlyAspPheLeuAsp 308  
235 CAGGGGAGTTTCTTTAGAC 252

seq\_name: gb\_sts:G71929

seq\_documentation\_block:

LOCUS G71929 312 bp DNA linear STS 08-JUN-2001  
DEFINITION A94818134FM017 maize leaf DNA Zea mays STS genomic, sequence tagged  
site.  
ACCESSION G71929 GI:14333614  
VERSION G71929.1  
KEYWORDS STS.  
SOURCE Zea mays.  
ORGANISM Zea mays.

REFERENCE 1 (bases 1 to 312)  
AUTHORS Yang,Y.J., Guo,L., Ashlock,D.A., Wen,T.J. and Schnable,P.S.  
TITLE 3' UTR sequences of maize genes  
JOURNAL Unpublished  
COMMENT

Contact: Schnable, P.S.

Schnable laboratory

Iowa State University

GA05 Agronomy Hall, Ames, IA 50011, USA

Tel: 515-294-0975

Fax: 515-294-2299

Email: schnable@iastate.edu

Primer A: GGAGACCAACCGTAGC

Primer B: TTAGACCAACCTGAGAGG

PCR Profile:

Denaturation: 94 degrees C for 30 seconds

Annealing: 60 degrees C for 45 seconds

Polymerization: 72 degrees C for 90 seconds

PCR cycles: 31

Thermal cycler: Perkin Elmer TC

Protocol: Template: 10-20 ng

Primer: each 0.5 uM  
 dNTPs: each 200 uM  
 Taq Polymerase: 0.05 units/uL  
 Total vol: 20 uL

## Buffer:

MgCl<sub>2</sub>: 2 mM  
 KCl: 50 mM  
 Tris-HCl: 20 mM  
 pH: 8.4

FEATURES  
 source

1. 312  
 /organism="zea mays"  
 /strain="DEB11"  
 /db\_xref="taxon:4577"  
 /clone\_lib="maize leaf DNA"  
 /note="PCR products amplified from genomic DNA"

STS  
 BASE COUNT 86 a 82 c 54 g 90 t  
 ORIGIN

## alignment\_scores:

Quality: 54.00 Length: 54  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-805-550-4 x G71929/rev ..

Align seg 1/1 to reverse of: G71929 from: 1 to: 312

315 MetProHisAlaIleSerValThrProGluGluGluGluAlaIleGly 331  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 311 ATGCGTCATCGCATCTAGTGTACACAGAGAGAGAGAGCGCATTTGACG 262  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 331 gtleuglsermetgllypheaspargalaaargvalillegluAlaIle 348  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 261 GCTTGAGTCCATGGGTTTCAGACAGACGCGTTATCGAAGCATTCCTAG 212  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 348 lacysaspargasngluGluGluAlaIleAsnTyrLeuGluHisAla 364  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 211 CCGCGCATAGAGACGAGAGAGCTAGACCAACTCTCTCTTGAGCATGCT 162  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 365 GlyGluGluAsp 368  
 ||||||||||||||  
 161 GGTGAGAGAGAT 150

seq\_name: gb\_htg:AP004214

## seq\_documentation\_block:

LOCUS AP004214 70957 bp DNA linear HTG 27-SEP-2001  
 DEFINITION Oryza sativa chromosome 2 clone OJ1118\_B07, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, in ordered pieces.

ACCESSION AP004214  
 VERSION AP004214.1 GI:15787871  
 KEYWORDS HTG, PHASE2.  
 SOURCE Oryza sativa (cultivar: Nipponbare) DNA, clone: OJ1118\_B07.  
 ORGANISM Oryza sativa

REFERENCE  
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.  
 TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC  
 clone: OJ1118\_B07

REFERENCE  
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.  
 TITLE Published Only in Database (2001) In press

JOURNAL  
 REFERENCE  
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-SEP-2001) Takuji Sasaki, National Institute of  
 Agricultural Sciences, Rice Genome Research Program, Kannondai  
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan

## COMMENT

(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,  
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
 The nucleotide sequence of this BAC clone was generated by  
 combining Monsanto and RGP-Japan sequencing data.  
 NOTE: It currently consists of 1 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces is believed  
 to be correct as given, however the sizes of the gaps between them  
 are based on estimates that have provided by the submitter. This  
 sequence will be replaced by the finished sequence as soon as it is  
 available and the accession number will be preserved.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

FEATURES  
 source

1. 70957  
 /organism="Oryza sativa"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /chromosome="2"  
 /clone="OJ1118\_B07"

BASE COUNT 20235 a 14617 c 14676 g 21429 t  
 ORIGIN

## alignment\_scores:

Quality: 23.00 Length: 23  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-805-550-4 x AP004214 ..

Align seg 1/1 to: AP004214 from: 1 to: 70957

163 AsplysValGlnArgAlaLeuArgAlaIleTyrAsnAsnProGluAlaGly 179  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 46582 GATAAAGTCACCAAGAGCTCTACGTCAGCTTAACAACCCAGACGGTGC 46631  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 179 aValGluTyrLeuTyrSer 185  
 ||||||||||||||  
 46632 TGTGCAATATCTATATTCT 46650

seq\_name: gb\_pl:AY063103

## seq\_documentation\_block:

LOCUS AY063103 1129 bp mRNA linear PLN 26-NOV-2001  
 DEFINITION Arabidopsis thaliana putative RAD23 protein (F20B17.8) mRNA,  
 complete cds.

ACCESSION AY063103  
 VERSION AY063103.1 GI:17104776  
 KEYWORDS FLI CDNA.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana

REFERENCE  
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 1129)

REFERENCE  
 AUTHORS Yamada, K., Ban, J., Banno, F., Chang, E., Dale, J.M., Goldsmith, A.D.,  
 Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C., Toriumi, M., Wu, H.C.,  
 Yamamura, Y., Yu, G., Yu, S., Bower, L., Carninci, P., Chen, H.,  
 Cheun, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A.,  
 Karlin-Neuman, G., Kawai, J., Kim, C., Koese, E., Lam, B., Lin, J.,  
 Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J.,  
 Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A.,  
 Shinzaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.  
 Full Length cDNA of gene F20B17.8 (GI:1715605)

REFERENCE  
 AUTHORS Yamada, K., Ban, J., Banno, F., Chang, E., Dale, J.M., Goldsmith, A.D.,  
 Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C.,  
 Yamamura, Y., Yu, G., Yu, S., Bower, L., Carninci, P., Chen, H.,  
 Cheun, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A.,  
 Karlin-Neuman, G., Kawai, J., Kim, C., Koese, E., Lam, B., Lin, J.,  
 Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J.,  
 Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A.,  
 Shinzaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.  
 Full Length cDNA of gene F20B17.8 (GI:1715605)

JOURNAL  
 REFERENCE  
 AUTHORS Yamada, K., Ban, J., Banno, F., Chang, E., Dale, J.M., Goldsmith, A.D.,  
 Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C.,  
 Yamamura, Y., Yu, G., Yu, S., Bower, L., Carninci, P., Chen, H.,  
 Cheun, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A.,  
 Karlin-Neuman, G., Kawai, J., Kim, C., Koese, E., Lam, B., Lin, J.,  
 Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J.,  
 Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A.,  
 Shinzaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.  
 Full Length cDNA of gene F20B17.8 (GI:1715605)

TITLE  
JOURNAL

## COMMENT

Karlin-Neumann, G., Kawai, J., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.  
Direct Submission  
Submitted (15-NOV-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA  
The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and sequenced the pPNT (ORF) clones using the RAPL cDNAs: Yamada, K., Banh, J., Banno, F., Chang, E., Dale, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shim, P., Southwick, A., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

FEATURES  
source

1..1129  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="U09913"  
/note="This clone is in pPNT 51.  
ecotype: Columbia"

## gene

1..1129  
/gene="F20B17.8"

## CDS

1..1098  
/gene="F20B17.8"  
/note="contains two UBA PF00627 domains and is a member of the ubiquitin PF00240 family"

/evidence="experimental"  
/product="putative RAD23 protein"  
/protein\_id="FAL34277.1"  
/db\_xref="GI:17104777"  
/translation="MKLVKTKLGSHEFIVLPSDTMAVKNIEDSOGKDNYPGCGO  
LTIHNGKVLNDETSLEKNKTEEGFLVNLKSKSGSAOASQVPSATTSSTKPA  
PSTTQSSVPASPIPAQEQPAQDTGQAATLVSSSLEMOQOIMENGSGMSDE  
TTRALRAAANNPERAVDLYSGIPQAEVAVPEAIOIGSGAIPVAPASGGGNSNP  
LDLPEQETVAASGDLGTEFLRNDDFOOLRTMHSNPOIIOQMCELGKPNOLL  
RLIOENAEFLQALVNEPYEGSDGDFDPEQEMPHAINVTAEQZAIORLEAMGFD  
RALVIEAFIACDKREELIANTYLENSGDFED"

misc\_difference 576  
/gene="F20B17.8"  
/note="compared to genomic sequence"  
/replace="g"  
1099..1129  
/gene="F20B17.8"

BASE COUNT 332 a 254 c 261 g 282 t  
ORIGIN  
Alignment\_scores:  
Quality: 21.00 Length: 21  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment\_block:  
US-09-805-550-4 x AY063103 ..

Align seg 1/1 to: AY063103 from: 1 to: 1129

342 VALLIAGLUAhPheLeuAlaCysAPARGaNGIuLeuAlaAlas 358  
|||||

1012 GTAAATGAAAGCGTTCTCCGATGTGACCGAATGAAGAAATTGGCAGCTAA 1061  
358 nrryleuLeuGlu 362  
|||||

1062 CTATCTATTGGAG 1074  
seq\_name: gb\_pl:AY034912

seq\_documentation\_block:

LOCUS AY034912 1518 bp mRNA linear PLN 10-JUN-2001  
DEFINITION Arabidopsis thaliana putative RAD23 protein (F20B17.8) mRNA.

complete cds.

ACCESSION AY034912  
VERSION AY034912.1 GI:14334441  
KEYWORDS F11 cDNA.

## SOURCE

## ORGANISM

Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 1518)

## REFERENCE

## AUTHORS

Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K.,  
Goldsmith, A.D., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Yu, G.,  
Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,  
Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,  
Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M.,  
Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M.,  
Shim, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and  
Theologis, A.  
Full length cDNA of gene F20B17.8 (GI:7715605)  
Unpublished  
2 (bases 1 to 1518)

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K.,  
Goldsmith, A.D., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Yu, G.,  
Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,  
Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,  
Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M.,  
Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M.,  
Shim, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and  
Theologis, A.  
Direct Submission

Submitted (10-MAY-2001) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN  
Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J.,  
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the  
sequencing and annotation of the RAPL cDNAs: Yamada, K., Liu, S.X.,  
Sakano, H., Pham, P.K., Banh, J., Chung, M.K., Goldsmith, A.D.,  
Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Yu, G., Bowser, L.,  
Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E.,  
Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J.,  
Shim, P., Southwick, A., Davis, R.W., Ecker, J.R. and Theologis, A.  
Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to  
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)  
contributed equally to this work as PIs.

## FEATURES

## source

1..1518  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="RAF105-08-C09 (R09913)"  
/note="This clone is in a modified pBluescript vector  
(lambda zap) as a XhoI/SstI insert.  
ecotype: Columbia"  
1..136  
/gene="F20B17.8"

5' UTR  
gene

CDS

137..1234  
 /gene="F20B17.8"  
 /note="contains two UBA PF00627 domains and is a member of the ubiquitin PF00240 family"  
 /codon\_start=1  
 /evidence=experimental  
 /product="putative RAD23 protein"  
 /protein\_id="AAK59419.1"  
 /db\_xref="GI:14334442"  
 /translation="MKLVTKLKGSHFEIRVLPSTIMAVKNKIEDSGKNDYPCGOQLIINGKVIDKETSLVENKVEEGFLVWLSKSGSGAGASVQPVATTSSTPDAPSTOSSPVASPIPADOPAPOTDTYGOAASSTYSSGSECOMQIEMEGSSMDKE TTVTRALRAHNNPERANDYISGIPOTAEVAVPPPEAIASGAAPVAPASGPNSSP LDFPEQETVAAGSGDGTLEFLRNNDQFQQLRTVWHSNPQIIPMLQELCKNQPLL RLIDNQAEPFLQLVNEPEYSGSDGEMDFDEOEPMFPAIINVTPAEQEAIORLEAMGFD RALVTEAFVACDRNEELAAVNIENSGDFED"

misc\_difference  
 /gene="F20B17.8"  
 /note="compared to genomic sequence"  
 /replace="g"  
 3'UTR  
 1235..1518  
 /gene="F20B17.8"  
 misc\_difference  
 1502  
 /gene="F20B17.8"  
 /note="not present in genomic sequence"

BASE COUNT 421 a 326 c 338 g 433 t  
 ORIGIN

alignment\_scores:  
 Quality: 21.00 Length: 21  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-805-550-4 x AY034912 ..

Align seg 1/1 to: AY034912 from: 1 to: 1518

342 ValIleGluAlaPheLeuAlaCysAspArgAsnGluLeuLeuAlaIleAs 358  
 |||||  
 1148 GTATATGAACCGCTTCCTCGCATGTGACGAAATGMAAGAAATTGGCAGCTAA 1197  
 358 nTYrLeuLeuGlu 362  
 |||||  
 1198 CTATCTATTGGAG 1210

seq\_name: gb\_pl:AC010793

seq\_documentation\_block:  
 LOCUS AC010793 90149 bp DNA linear PLN 07-MAY-2000  
 DEFINITION Genomic sequence for Arabidopsis thaliana BAC F20B17 from chromosome I, complete sequence.  
 ACCESSION AC010793  
 VERSION AC010793.3 GI:6938146  
 KEYWORDS HTG.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 90149)  
 Khan,S., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Kim,C., Shinn,P., Altafi,H., Bel,O., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N., Howng,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,T., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Federspiel,N.A., Theologis,A. and Ecker,J.R.  
 TITLE  
 Genomic sequence for Arabidopsis thaliana BAC F20B17 from chromosome I  
 JOURNAL  
 Unpublished  
 REFERENCE  
 2 (bases 1 to 90149)

AUTHORS Ecker,J.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-SEP-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA  
 REFERENCE  
 3 (bases 1 to 90149)  
 Ecker,J.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-FEB-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA  
 4 (bases 1 to 90149)  
 Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N., Howng,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,T., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-MAY-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018 USA  
 On Feb 8, 2000 this sequence version replaced gi:6453825.  
 FEATURES  
 source  
 1..90149  
 /organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"  
 /chromosome="1"  
 /clone="F20B17"  
 complement(join(3669..4322,4410..4681,5385..5574))  
 /note="hypothetical protein"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="F20B17.1"  
 /protein\_id="AAF68129.1"  
 /db\_xref="GI:7715611"  
 /translation="MEIGSSSTVAGGGLSVPPGFRFRPTREELLYYLKKVSEPI DUDVIREVDLNLKEPELKEKCRIGSGFQNEWYFYSKDKKYPTGTPTNRATAGFWK ATGRDSSIHNSKRIGLRKTLYFTYGAPRQKQKHEWHEHRLDSSNEISLEQGVVY CRVFKKNHPRGFQRODHHNHHYQISTNDDHHHHHIIHDSNNSPILILHPRDHH HHHHNGOIRHPRLEPANTLSHGSMLRPLEPSPSAAAAAASAPFPYSPITTD IECSONLRLTISNNNYGDWFLDKLTLTGNNDQDQDQVQNHQAKCGDJSNDND QADHIGNNGSSSSPVNQREFHYHGDANLKRPK"  
 join(9713..9930,10050..10255,10351..10577,10822..10953)  
 /note="unknown; similar to ESTs gb|T04661.1, and emb|235002.1"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="F20B17.2"  
 /protein\_id="AAF68106.1"  
 /db\_xref="GI:7715588"  
 /translation="MASSSDPMWRREYNALIKSIDINGNMSRNASGITGPDACRRAS ATRKTTILGTRLDLSLQSLVAVPGKHVSEKERRKRDVGNLSKTNQVASALNMS NFANDSELTGTLKRPDAINRVSGMDNGIVFOQVAKRDEDEGLEKLEYVMSTKHI ALAVNEFLGTRLIVCLRCSCILFLILISVSNMGSSMDDDIDYVDITDSRLRR VOKSLAMNKSMSKSCSCMSMLSVGIYGLAVIWLKTYL"  
 complement(join(12754..12882,12967..14412,14530..14757, 14833..15165))  
 /note="unknown; similar to ESTs emb|230470.1, gb|A193698.1, gb|N96085.1, and emb|230471.1"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="F20B17.3"  
 /protein\_id="AAF68128.1"  
 /db\_xref="GI:7715610"  
 /translation="MSLVVGGSLGLTVGDLSLNSKTIWKSFKFSVNRRAAA LVOAPRKEDGAASPSPSPASVVOYRRADLADLOAEARLGAIDASISPELIA RKHGGSPKALRSLRLLEIGALGFAKGIQIOKQNLKNNMKRAIEIRLFTPTGPT FVKLGOGSTRPDLCPPYLELAELOALPPEAPAEACIERLDSLETFPSVS PPPIAASLGOYVKAOLRYSGOVAVVAKVORCIEAIDLEFLIINGVCKLKNKYVFI TTDTITLIDERRACRYOELINVQOENARREKRLADADVAVPIENVYSRVKVTM FWEGTKLNEOLAIESGILKVLIVNTIGICSLRQLLEYGFFHADPHGNNLAPDGG

CDS

LAFLDRGMSSEPEAFALIGHVHLVNDYEMANDYALAKFLSPDVTPIPL  
RDFPDDALNTVSELNFKTLVDGLGAVFYQFENVRYALILKSLTVEGLADYAP  
NKFVLAASYPAKRLITDNPYLDLILFELDFGRNRLNLELLOQSGKSLADYAP  
KDALQPLKLLDNPNGEELRLVKEAVRSEALCTVDYDITLPEFLRSYFNN  
GNGPLTWSTAELOSTLELDROVSRIMGLVSSSEFPAIIQPIQVLOQEPARLGR  
VAGVGORLARFLQOLLRATTPSSASP"

complement(join(15862, .15982, 16100, .16200, 16324, .16407,  
16541, .16832, 17100, .17144, 17245, .17342, 17518, .17678,  
18130, .18224, 18308, .18384, 18435, .18497, 18587, .18698,  
19119, .19194, 19351, .19434, 19533, .19719))  
/note="similar to sodium-hydrogen exchanger NHE1  
(AF142676)"

CDS

/codon\_start=1  
/evidence=not\_experimental  
/product="F20B17.4"  
/protein\_id="AA68127.1"  
/db\_xref="GI:7715609"  
/translation="MSSELOISPAIHDPGQKQQAAGVILLOIMLVISFLVGHV  
LRHKFYLLPEASNLIGLVGLANISNTERTSIRVEFLISFRHSGISTSSSF  
CFCLPSTYILKLEYLGVMFLRLDFEVLCEMLGSLISADPTVLSTIQELSDVN  
LVALVESEVSLANDMEIVTLIRSFSLCQWMAISLYRTMSLVRSMSGONFVVI  
VRLPEVGSMSAMKFIILMYSLLSVTRTWSAVSSYFPHSHKTLTYTSVSYI  
FTLEIVQFMKHTYSNLSANSOREFAEFHLISLAETFEVFIYMGEVLAEMESHA  
ANYEGGCIYVNLARPAHRTDPTHOKALMYSGKTLICVPLSCYFSSVNTKCGCC  
IGLRGMAFALALQSVHDEPEGHQITFTATTAIVTVLLIGSTMTALEVVD  
SHDTSLDGEVNVSRITSYDDEDPGSGFRKLEFRHSAASFELDRNTLTPFE  
TSNNGDYDEGNMEOHGNHNL"

complement(join(20282, .20788, 20865, .21100, 21175, .21394,  
21491, .21630, 21764, .21937, 22013, .22701, 22782, .22853,  
22939, .23013, 23095, .23163, 23251, .23322, 23560, .23634,  
23717, .23878, 23973, .24044, 24146, .24220, 24312, .24383,  
24506, .24629, 24713, .24823))  
/note="unknown; similar to Ests gb|A197133.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F20B17.5"  
/protein\_id="AA68126.1"  
/db\_xref="GI:7715608"

CDS

/translation="WGSNTVATSRLLICAFASFTVFSNMISSVTPDRDAALRLSLM  
DNDNTPPSGSGDDPCGTPEGVSCNNSHITAGLSTMGKRLSGDIELAELRL  
DLSPNRLTGLSLRLDLOKRLNLLIAGSGFTGTIPNEGLKDLSPALNSNFPD  
KIPASGLNTRKWLADLADNOLNGPIRISGSSPGDLLLAKRHFHNSKLGSTIP  
KLFSSEMILIHVRLDRNTLTKVPEPENSNTJNTIELANIKLHFSKJLSDMSMAY  
VLSNNSFDPSESPLEPSTPLTLYMEGSLDGPPLNKLFGPQLOQVRLKNAAN  
GLSLDGVPELQVLDLQNDISSVLSGYNTLWYIASFELNVSOLCYQDITKRA  
LSCRLGPNVCTALSNNTYCOIQOQVARIYSTLANCGKSCDQVSPSCC  
AYVEGTLYRGPMPFRLSNVNTYHSLMSLWGLTLPGSVSLQNPFNNDYLOIO  
LALFPPKGYENREYORIGEDLSNOTKPPLEGPYFTASPTTPADNGHLSR  
MNGITGCSALVIALGITYAMMOKRRAEOATIGSKYKMLDGHVATIRAOGST  
GAPOUKARWFSTYELKKTITNNTSVSELSLGTGKTKMLDGHVATIRAOGST  
OGLEFTELELSRVHKNLVGLVGCFFDQEQILVLEYMSNGLSDSLTRGKITL  
DMKRLVALAGSARGLAYLHLDAPPIIHSDVYSTNLIDENLTAKADGSLKLVSD  
CTKGHVSQVKGTLGYLDPEYTTQKLTKESDVYSFVVMMLTAKOPIEKGYIVR  
EIKLYMNKSDPDFGLRDKMDSLVDGTLPELGRVYELLAKCVDETADRPPTSEVY  
KETEIIIONGASASSSSASASSATDGEHLVGGITAKKEARDGDCGAFDISGGS  
VPTKIEPK"

complement(join(25580, .26059, 26154, .26582, 26669, .27097,  
27555, .27818))  
/note="hypothetical protein"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F20B17.6"  
/protein\_id="AA68125.1"  
/db\_xref="GI:7715607"  
/translation="MGLCYSVDRITGKEPEASATTAETVEERSGSGRRMRPDK  
GGDIBGIPOLVGLVNSGSKIACLTQOGKSGSSNDVANGNCRDQYVTCGAF  
DGHGPRGHMVAKRYDRLPTTLTQLKMTESQSSSYVANGROIKTEEBEYVOTTS  
EQVKTESVTHDEQCELPNPNNDDELPMITLPLKAMLSQOQIDKEIKMHTTDC  
FCSGTTSVTLIKOGEDELVGNIDSRVAVLTREDNALAVQITLIDKPLGLTLOI  
MINSERSQOKELNFRFLCVLGSSEARVILKCGRVALLDEPVAVMJLPSNSPGL  
AMARAFDCLKILSVDPIDINVRILTEDDOFLIADSDVAVLSKRVKVDIYASAP  
SRSTAAALVDTAVRSWIRIKPTSKNDKCVCLFDOSVAVMSVNVKDKDEES  
IESVTNISTKEDEIYVDEKIPESGIESKMTATLACISVADDEERSALEGIT

CDS

RVNSLSIPRLSGELNRSWKRL"  
complement(join(29562, .29631, 29861, .29925, 30172, .30291,  
30565, .30659, 30757, .30869, 31083, .31133, 31431, .31498,  
31643, .31706, 31868, .31893, 32190, .32360, 32589, .32670,  
32937, .33061, 33213, .33467, 33581, .33739))  
/note="hypothetical protein"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F20B17.7"  
/protein\_id="AA68124.1"  
/db\_xref="GI:7715606"  
/translation="MKKKYPIDGHEHYTYEFGQGSALVHRLCPDPVVAIKTL  
DPERNCDLNNISREAOITMLVDPNPLKSHCSFVSHNIMVPIPHSGSCCHITLKA  
AYPDGEFEALIAITLREALGLDYLHGHGHIHRVYKGNLLILGARVAKLGDVGSVC  
LFDSGDRQRTNRTFVGTCPNADWISFETIGELIAGHAPSPSKYPPKSPFOMIASCLV  
KDPKRSRPAKLLKHSFKOARSDDYIARLKLGLDPLVNRVOAIKNEYKRGISGLSG  
SAYSLALDSDODOSEIOEDPGITNLYLOPLIHRSLIARSDSDSDSLASPSYDY  
VYSPRHEDSLNTHYSGVHANNKGTQDTSPTNPTETINGSVSLAONKAPNG  
ESDTQOQLONGSNCKNTHPTVGSDVPTEMAYKPPKASSLSDSKKPPVQORG  
RFRKTSNLDIEKVLCOHSSASLPHSDVTLPLNLTSSVYLPVLPVLTNLERDVOI  
NKEVYNNILCPOLRNPMWLICAEELQKTEHQAQVSMN"

complement(join(35055, .35171, 35264, .35348, 35471, .35580,  
35968, .36075, 36158, .36283, 36360, .36530, 36765, .36828,  
36929, .36984, 37041, .37079, 37300, .37461, 37716, .37811))  
/note="similar to RAD23, isoform I (F12013); similar to  
Ests gb|T42959.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F20B17.8"  
/protein\_id="AA68123.1"  
/db\_xref="GI:7715605"

CDS

alignment\_scores:  
Quality: 21.00 Length: 21  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment block:

US-09-805-550-4 x AC010793/rev ..

Align seg 1/1 to reverse of: AC010793 from: 1 to: 90149

342 ValIIEGLuAlaPheLeuAlaCyAsPaTgaSngLugLLeuAlaLAs 358  
|||||  
35141 GAAATAGAAAGCCCTTCCTCGCATGTGACCAAAATGAAGATTGGACGCTAA 35092

358 nTyrrLeuLeuGlu 362  
|||||

35091 CTATCTATTGGAG 35079

seg\_name: gb\_htg:AP003881

seg\_documentation\_block: 110775 bp DNA linear HTG 10-JUL-2001

LOCUS AP003881 110775 bp DNA linear HTG 10-JUL-2001  
DEFINITION Oryza sativa chromosome 8 clone OJ1124\_B05, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, in ordered pieces.

ACCESSION AP003881  
VERSION AP003881.1 GI:14646814  
KEYWORDS HTG; HTGS PHASE2.  
SOURCE Oryza sativa (cultivar: Nipponbare) DNA, clone: OJ1124\_B05.  
ORGANISM Oryza sativa

REFERENCE 1 (bases 1 to 110775)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
TITLE Oryza sativa nippobare (Ga3) genomic DNA, chromosome 8, BAC  
clone: OJ1124\_B05

JOURNAL 2 (bases 1 to 110775)  
published only in Database (2001) in press  
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.  
TITLE Direct Submission

JOURNAL Submitted (09-JUL-2001) Takuji Sasaki, National Institute of

Agrobiological Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsasaki@abrr.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data.

NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

#### FEATURES

##### source

1..110775  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/chromosome="8"  
/clone="OJ1124.B05"

BASE COUNT 31727 a 23364 c 23821 g 31760 t 103 others  
ORIGIN

#### alignment\_scores:

Quality: 18.00 Length: 18  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-805-550-4 x AP003881/rev ..

Align seg 1/1 to reverse of: AP003881 from: 1 to: 110775

167 ArgAlaLeuArgAlaAlaTyrAsnAsnProGluArgAlaValGluTyrLe 183  
|||||  
86919 CGAGCTCTGGAGCTGCATACACATCCAGAACGGCGCTGGAGTATT 86870

183 uTyr 184  
|||||

86869 GTAC 86866

seq\_name: gb\_pl:AY037181

#### seq\_documentation\_block:

LOCUS AY037181 1378 bp mRNA linear PLN 10-JUN-2001  
DEFINITION Arabidopsis thaliana unknown protein mRNA, complete cds.  
ACCESSION AY037181  
VERSION AY037181.1 GI:14335003

#### KEYWORDS

##### SOURCE

#### ORGANISM

thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

#### REFERENCE

##### AUTHORS

#### TITLE

#### JOURNAL

2 (bases 1 to 1378)  
Unpublished  
Cheuk, R., Chen, H., Kim, C.-J., Koesema, E., Meyers, M.C., Shinn, P.,  
Tracy, S.E., Ban, H., Bowser, L., Carninci, P., Jones, T., Kamiya, A.,  
Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Lee, J.M., Lin, J., Liu, S.X.,  
Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Pham, P.K.,  
Quach, H.L., Sakano, H., Sakurai, T., Satou, M., Seki, M., Southwick, A.,  
Tang, C.C., Tortum, M., Yamada, K., Yu, G., Shinozaki, K., Davis, R.W.,  
Theologis, A. and Ecker, J.R.  
Arabidopsis cDNA clones

#### TITLE

##### JOURNAL

Karlín-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X.,  
Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Pham, P.K.,  
Quach, H.L., Sakano, H., Sakurai, T., Satou, M., Seki, M., Southwick, A.,  
Tang, C.C., Tortum, M., Yamada, K., Yu, G., Shinozaki, K., Davis, R.W.,  
Theologis, A. and Ecker, J.R.

Direct Submission  
Submitted (22-MAY-2001) Salk Institute Genomic Analysis Laboratory  
(Signal), Plant Biology Laboratory, The Salk Institute for  
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,  
USA

#### COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RFL cDNAs (RFL cDNA : RIKEN  
Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J.,  
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PEGC (SSP) Consortium members carried out the  
sequencing and annotation of the RFL cDNAs: Cheuk, R., Chen, H.,  
Kim, C.J., Koesema, E., Meyers, M.C., Shinn, P., Tracy, S.E., Ban, H.,  
Bowser, L., Chung, M.K., Goldsmith, A.D., Jones, T., Karlín-Neumann, G.,  
Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M.,  
Palm, C.J., Pham, P.K., Quach, H.L., Sakano, H., Southwick, A.,  
Tang, C.C., Tortum, M., Yamada, K., Yu, G., Davis, R.W., Theologis, A.,  
and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to  
this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)  
contributed equally to this work as PIs.

#### FEATURES

##### source

1..1378  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="RFL09-12-L08(R12657)"  
/note="ecotype: Columbia"  
1..70  
71..1207

#### 5'UTR

##### CDS

/codon\_start=1  
/product="unknown protein"  
/protein\_id="AAK59766.1"  
/db\_xref="GI:14335004"  
/translation="MKIFVKTLSGSNFELEVKPADKVSQVKALETWKAEPAAKQK  
LIHQKVLKDELTENNVENSTFVIMLSKTAKSPGSTASAPASAPQVATP  
QVSAPASVPEPTSGTAAAPATAASVDTQVGAQSMVAGTLESTVQOILMGCG  
GSMRDRTVRLAAAFNPERAVEYLSGIPQAEIPPAQVPAQGEAANLAPQO  
AAAPAAATGGPNANPLNFERPQGPADAGAGNLDLFINSHQPAIRAMVANGAOIT  
QPMQDELGRNPQVRLIOEHADFLRLINEVEEGENMEQLEAAMPQAVVTPEER  
DAIERLEGGFQRAVYLEVFACNKEELANLIDHMEFEQO"  
1208..1378

#### BASE COUNT

3'UTR 366 a 332 c 296 g 384 t

#### alignment\_scores:

Quality: 15.00 Length: 15  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-805-550-4 x AY037181 ..

Align seg 1/1 to: AY037181 from: 1 to: 1378

174 AsnAsnProGluArgAlaValGluTyrLeuTyrSerGlyIlePro 188

|||||  
599 AACATCTCTGAAGAGCTGTGAATACCTGTACTGGGATATCCCT 643

seq\_name: gb\_pl:AY058196

#### seq\_documentation\_block:

LOCUS AY058196 1426 bp mRNA linear PLN 04-NOV-2001  
DEFINITION Arabidopsis thaliana unknown protein mRNA, complete cds.  
ACCESSION AY058196



VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

AY058196.1 GI:16648837  
FLI\_CDNA.  
thale cress.  
Arabidopsis thaliana  
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 1426)  
Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J.,  
Bowser,L., Carninci,P., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,  
Ishida,J., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,  
Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,  
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K.,  
Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,  
Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,  
Davis,R.W., Theologis,A. and Ecker,J.R.  
Arabidopsis cDNA clones  
Unpublished  
2 (bases 1 to 1426)  
Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J.,  
Bowser,L., Carninci,P., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,  
Ishida,J., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,  
Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,  
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K.,  
Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,  
Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,  
Davis,R.W., Theologis,A. and Ecker,J.R.  
Direct Submission  
Submitted (28-AUG-2001) Salk Institute Genomic Analysis Laboratory  
(SIGAL), Plant Biology Laboratory, The Salk Institute for  
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,  
USA  
RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN  
Arabidopsis Full-Length CDNA') : Seki,M., Narusaka,M., Ishida,J.,  
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,  
Hayashizaki,Y. and Shinozaki,K.

COMMENT

The Salk, Stanford, PGEC (SSP) Consortium members carried out the  
sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H.,  
Kim,C.J., Koesema,E., Meyers,M.C., Shin,P., Banh,J., Bowser,L.,  
Dale,J.M., Goldsmith,A.D., Jiang,P.X., Jones,T., Karlin-Neumann,G.,  
Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M.,  
Onodera,C.S., Palm,C.J., Pham,P.K., Quach,H.L., Southwick,A.,  
Tang,C.C., Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S.,  
Davis,R.W., Theologis,A. and Ecker,J.R.

FEATURES

Source

1. 1426  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="RAFL09-10-108(R14049)"  
/note="ecotype: Columbia"  
1. 78  
79. 1215  
/codon\_start=1  
/product="unknown protein"  
/protein\_id="AA125609.1"  
/db\_xref="GI:16648838"  
/translation="MKLFVTLTSGSNEIEVKKPADKSDVKTALTEVKAEPRAKQK  
LIHQKVLKDETLLENVNVSEIVIMLKTRKSPGASTASAPASATOPOTVATP  
QVAPFASVPTSGATTAAPATPAASGVDTGOAASNNVAGTTLESTVOILDMG  
GSDRDVTVVATRAAFNNPRAVEYLSGIPDAEIPVPAAGTGOAANPAAPOO  
AAAPAAATGCPNANPLNLFPGCPADAGAGAGUDELRSOQFALRAVQANPQIL  
QPMQLGKQNPOLVLIQEHQADFRLINPEVGEENVEQLEAANPQATVTPPER  
EAIERLEGMGFDRAMVLEVFACKNKNEELANVLLDHMEFDD"

5' UTR  
CDS  
3' UTR  
BASE COUNT  
ORIGIN

377 a 343 c 301 g 405 t

alignment\_scores:  
Quality: 15.00 Length: 15  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-805-550-4 x AY058196 ..  
Align seg 1/1 to: AY058196 from: 1 to: 1426

174 AaasnpProGUargAlaValGufYrLeuYrSerGlyIlePro 188  
|||||  
607 AACAAATCTGAAGAGCTGTGAATACCTGACTCGGATCCCT 651

seq\_name: gb\_pl:DCRAD2311

seq\_documentation\_block:  
LOCUS DCRAD2311 1466 bp mRNA linear PLN 28-JAN-1999  
DEFINITION Daucus carota mRNA for RAD23 protein, isoform II.  
ACCESSION Y12014  
VERSION Y12014.1 GI:1914684  
KEYWORDS isoform II; RAD23 (ubiquitin-like protein involved in UV excision repair).  
SOURCE carrot.  
ORGANISM Daucus carota  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; eusterids II; Apiales; Apiaceae; Daucus.  
REFERENCE 1 (bases 1 to 1466)  
AUTHORS Sturm,A. and Lienhard,S.  
TITLE Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in yeast  
JOURNAL Plant J. 13 (6), 815-821 (1998)  
MEDLINE 98345997  
REFERENCE 2 (bases 1 to 1466)  
AUTHORS Sturm,A.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-1997) A. Sturm, Friedrich Miescher-Institut, Maulbeerallee 66, CH-4058 Basel, SWITZERLAND  
FEATURES  
Source

1. 1466  
/organism="Daucus carota"  
/isolate="W001C"  
/sub-species="Queen Anne's Lace"  
/db\_xref="taxon:4039"  
40. 1179  
/function="assembly factor of the complex for nucleotide excision repair of V-damaged DNA"  
/codon\_start=1  
/product="RAD23 protein, isoform II"  
/protein\_id="CAA72742.1"  
/db\_xref="GI:1914685"  
/db\_xref="SPRMBL:003991"  
/translation="MKLVTKGSHPEIRAPNDYMAIKKNIEDLQGRDNPCGQO  
LLIHNGVLYKDESTLAESKISEDFLVLMKSKTMSSTGTPAQSASAPAPAPAV  
APAPAPAPASAVIPNTTVPEAPISFAFASIDYVGAASNVVAGSNLEGTIHKM  
GGGMQDTNVSRAIRAYNNPERAVDLYSGIPMAEAVVSHFGQDOINAGNNATS  
DNGVGAAPAGAPNSLPINMFPEOFTLSCVTGAGCSLEFLNNPQFRLRBMVRNPQI  
LQPMLELQKQNPOLLQIOEHHEEFLQILNERPEVASEGCMFDDPEDDVYQETIVTA  
DQEAIERLEAMGDFRGVITFAFACDNEELANVLLNAGDFDD"

CDS  
BASE COUNT  
ORIGIN

405 a 311 c 326 g 424 t

alignment\_scores:  
Quality: 15.00 Length: 15  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-805-550-4 x DCRAD2311 ..



Align seg 1/1 to: DCRAD231I from: 1 to: 1466

342 ValIleGluAlaPheUeAlaCysAspArgAsnGluGluLeuAla 356  
 |||||  
 1093 GTTATTGAGACCTTTTACCATCTGCACCCGACGAGGAATTGGCG 1137

seq\_name: gb\_pl:DCRAD231

seq\_documentation\_block:

LOCUS DCRAD231 1496 bp mRNA linear PLN 28-JAN-1999  
 DEFINITION Daucus carota mRNA for RAD23 protein, isoform I.  
 ACCESSION Y12013  
 Y12013.1 GI:1914682  
 KEYWORDS isoform I; RAD23 (ubiquitin-like protein involved in UV excision repair).  
 SOURCE carrot.  
 ORGANISM Daucus carota  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; easterids II; Apiales; Apiaceae; Daucus.

REFERENCE 1 (bases 1 to 1496)  
 Sturm, A. and Lienhard, S.  
 Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in yeast  
 Plant J. 13 (6), 815-821 (1998)  
 MEDLINE 98345397  
 REFERENCE 2 (bases 1 to 1496)  
 Sturm, A.  
 Direct Submission  
 Submitted (24-MAR-1997) A. Sturm, Friedrich Miescher-Institut, Maulbeerallee 66, CH-4058 Basel, SWITZERLAND  
 LOCATION/Qualifiers  
 1. 1496  
 /organism="Daucus carota"  
 /isolate="W001C"  
 /sub-species="Queen Anne's Lace"  
 /db\_xref="taxon:4039"  
 63. 1211  
 /function="assembly factor of the complex for nucleotide excision repair of V-damaged DNA"  
 /codon\_start=1  
 /product="RAD23, isoform I"  
 /protein\_id="CAI72741.1"  
 /db\_xref="GI:1914683"  
 /db\_xref="SPTREMBL:O03990"  
 /translation="MKIYVKTLKSGSEIOYVNPDSVAVKRSIETAGAAVYPAAO MLITQKYLKNGTTLLENVAVENSTIVIMLSKSPSEGSTTSTRAAPKAPQTSAPP SVPAVAVSQPASTLPVPAPSPAPAPATAPLPSAVGSEAVYDASLVAAGSLWEG AIQDILDMGGGTWDRDVIKIVRAAFNNPERAVEXLYSGIPEQAEAPAPSPGQA ANPLDPPAAAPAPASAGPNANPLDLEPQGLPDMGSAAGAGNIDFLRTNQFOALR AMVQSNPQIILQPMLOELGKNPHLMRLTLEQHOADFLQLINEMEGEGMLGHGQALR VTPERPRAIERLEAMGPFRELVLVEFACNKNNEELAAVYLLDHMEFE"

FEATURES

source

CDS

BASE COUNT 403 a 361 c 331 g 401 t

ORIGIN

alignment\_scores:  
 Quality: 15.00 Length: 15  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-805-550-4 x DCRAD231

Align seg 1/1 to: DCRAD231 from: 1 to: 1496

174 AsnAsnProGluAlaGluAlaGluTyrSerGlyLeuPro 188  
 |||||  
 621 AACATTCAGAGAGAGCTGTGAGTATCTTATTCGGAATCCCT 665

seq\_name: gb\_pl:F309

seq\_documentation\_block:

LOCUS F309 114498 bp DNA linear PLN 02-JUN-1999  
 DEFINITION Arabidopsis thaliana chromosome 1 BAC F309 sequence, complete  
 ACCESSION AC006341  
 AC006341.2 GI:4887257  
 KEYWORDS HTG.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 114498)  
 Vysotskaia, V.S., Schwartz, J., Yu, G., Toriumi, M., Lenz, C., Liu, S., Lee, J., Li, J., Kremetskaia, I., Liu, A., Luros, J., Gonzalez, A., Altati, H., Araujo, R., Chao, O., Conn, L., Conway, A.B., Dunn, P., Hansen, N., Huizar, L., Kim, C., Palm, C., Rowley, D., Shin, P., Walker, M., Davis, R.W., Ecker, J.R., Federspiel, N.A. and Theologis, A.  
 The sequence of BAC F309 from Arabidopsis thaliana chromosome 1  
 Unpublished (1999)  
 2 (bases 1 to 114498)  
 Theologis, A.  
 Direct Submission  
 Submitted (11-JAN-1999) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA  
 3 (bases 1 to 114498)  
 Theologis, A.  
 Direct Submission  
 Submitted (25-MAY-1999) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA  
 4 (bases 1 to 114498)  
 Theologis, A.  
 Direct Submission  
 Submitted (02-JUN-1999) Plant Gene Expression Center, 800 Buchanan St., Albany, CA 94710, USA  
 On May 25, 1999 this sequence version replaced gi:4139327.  
 The sequence of BAC F309 from Arabidopsis thaliana chromosome 1.  
 LOCATION/Qualifiers  
 1. 114498  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="1"  
 /clone="F309"  
 246. 901  
 /gene="F309.1"  
 join(246. 275. 353. 462. 608. 692. 785. 901)  
 /gene="F309.1"  
 /note="Similar to gb|Y12014 RAD23 protein isoform II from Daucus carota. This gene is probably cut off. Est  
 gb|AA651284 comes from this gene."  
 /codon\_start=1  
 /evidence-not\_experimental  
 /protein\_id="AAD34676.1"  
 /db\_xref="GI:4966345"  
 /translation="MVNSNPQIILQPMLOELGKNPDLRLIQENQAEFLQLNEPYEG SDGDVDFEOPQPEMPHSVYVTPPEQSEIERLEAGFDRATVTEAFILSDRNEELAAV YLEHSADEPED"  
 complement(1361. 2444)  
 /gene="F309.2"  
 complement(join(1361. 1387. 1506. 1625. 1899. 2444))  
 /gene="F309.2"  
 /note="ESTs gb|T04357 and gb|AA595092 come from this gene."  
 /codon\_start=1  
 /evidence-not\_experimental  
 /protein\_id="AAD34673.1"  
 /db\_xref="GI:4966342"  
 /translation="MGLNSKAEVAKSRKNAEAEOKDRQTRKEBOYREAGPSPSKA VKREBEAEKKAETAKLKEARLKBQEKELKLPKDKANVTYVPEVTEAEI IRRREDOVALAKKAEKSKKOTRKAAGEDEYKMTLVYTNDDSDSIEHHTYDEALAR IIVSDNLPVDRHPEKRLKASFRAVEVELPRLKSEPGILTHYQYDLTKMKKSPDN PLNQAAAAAANE"

FEATURES

source

gene

CDS

```

gene
3662..5562
/ gene="F309.3"
/ join(3662..5562)
/ gene="F309.3"
/ note="is a member of Pf00481 Protein phosphatase 2C family."
/ codon_start=1
/ evidence-not_experimental
/ protein_id="A034674.1"
/ db_xref="GI:4966343"
/ translation="MGICHSKIDKTRKENGATSTATTVEROSGRRLRRPRLDYSG
EISEIOVGRVNGSGSEIACLYTOOGKGTNODAMLVNENCSRDYVLCYFPOH
GPRGHVSKRVKRMPEFTSTOLKTTSGTROSSKNGINAPTCVDEQCEQLCEK
DEKLPPEMTPLKRALKTCQOMDKELKMPRTINCSCSTSTVVIKOGDLVNGNG
DSRAVALATRODANALVAVOLTIDLPKDPESARINHCGRPALODEPVARWMLRN
SDSPGLAMARAFEDPCLKIDYLSVDINRHLERDQYIILATDGVADYVNLKLEAD
IVASAPSDTAAVAVDOTAVRAMRLKPTNPTVPMETKEKEKPESCIEKKTTLAEC
HSHESEESTITSSKDADKEKASTNETVPMETKEKEKPESCIEKKTTLAEC
ISVKDEBMSALFGLTRVNSLSIPRFSELSNSSWRKWL"
/ complement(10192..11349)
/ gene="F309.4"
/ complement(join(10192..10323,10518..10562,10642..10742,
10834..11039,11135..11349))
/ gene="F309.4"
/ note="ESTs gb|F15498, gb|H37515, gb|T41906, gb|T22448,
gb|W43356 and gb|T20739 come from this gene."
/ evidence-not_experimental
/ codon_start=1
/ protein_id="A034675.1"
/ db_xref="GI:4966344"
/ translation="MASSSDSMWRAYNEAKLSEINGATISESSSNVTPDQORAS
AIRKLTIRGNKIDSLQSLAETHGKPRISKEKMRNRDYGRLRSKANOANALNMSN
FARSDILGPDIPDMSRVGTMDNGIVGQROVMREDEGLEDELEGTVMSTKHLA
LVAEEIDLOTRLIDLDYHVDYDSRLRRVQSLAVNMNMRSGSCMSMLLSVLGI
VGLAVYIWMLVKYM"
11967..13706
/ gene="F309.5"
/ join(11967..12044,12084..12224,12408..12954,13133..13706)
/ gene="F309.5"
/ note="Contains two Pf101344 Kelch motif domains."
/ evidence-not_experimental
/ codon_start=1
/ protein_id="A034677.1"
/ db_xref="GI:4966346"
/ translation="MPVSSVSLPCQNPFLSDFSVSIIVNGVSGKVINFEARFNC
SKASRRGSEDTPTISRVRGSLIPDLALRCLAKLSHGHYGLVECSRGMRDLVRA
DYSCYKARNGMSMLFLTERSKNOMVADPEADRMHPLPRTAVODGNHSGFACV
CVSNCLLVIGGCYAPSVSSPPIQKPYTKDMRDEPKQMKMVAISMRRPRTFACS
VSGKVVAGGRNLTHSRGIDSAEVDYDVAWRMELRAMRPOKDCSLSTRGFHYS
DOYGRABONSSEVFNPRDNTWSTVEDVWPPSRAMQRAVOVMKNDRYTTIVMGSBLIK
TRTDGEWYNGVSVSVLPNHPRELEAGYFPAALRNELVIGCVLWMBESGAGR
FDIVRLPVVACNPLDRPLMWRPKPCIPAGSIIIGCVLSLESSSP"
/ complement(14717..17027)
/ gene="F309.6"
/ complement(join(14717..15924,15993..16097,16178..17027))
/ gene="F309.6"
/ note="Similar to gb|A012423 wall-associated kinase 2
from Arabidopsis thaliana."
/ codon_start=1
/ evidence-not_experimental
/ protein_id="A034678.1"
/ db_xref="GI:4966347"
/ translation="MGVDARFLVLMILLRICGYAASFPLALRNCSDHGVNVPY
PFGIGCYKKNWFEIYCKSSSDQDPLILPRLRAVTSFLNLDPSISVYNKFYIOS
PLKHSCPNBDGYSSSLNKGSPFTSENNKTAVGCNKAAMNATGVLOICETFC
GNEIRSYKGNANTSCVYGKCCOMTIPPLLOVADVATEKREKOGVOVFLFOFITS
GSLFTPELMEYSEYTTIELMRLLDSYMSKRVLCNKNRFFEDSCSHNCEBGP
YIFGCGODIDECRDPHILNKCGKRCYANVUSICEKTPWPIILSGTSLSLILIFGM
LICKANRRKRVAAOKRRFQORNGLLQOOTSLFHSVNTKVFSSNDLENADIVFWM
SRILGOGGQGVYKMLDGMIVAVKSKALKLEENEEFINELIILSQIHNRVVKIL
GCLLETPILVEYFIPNRILFDHLNPSDEPFMSWEVEREICEVADVLSHSAVS
IPYHNDVSTNLTDEKHKRAKVSDFGISVAIDTHTLTITVOGTIGYVDPELQSN

```

```

gene
3662..5562
/ gene="F309.3"
/ join(3662..5562)
/ gene="F309.3"
/ note="is a member of Pf00481 Protein phosphatase 2C family."
/ codon_start=1
/ evidence-not_experimental
/ protein_id="A034674.1"
/ db_xref="GI:4966343"
/ translation="MGICHSKIDKTRKENGATSTATTVEROSGRRLRRPRLDYSG
EISEIOVGRVNGSGSEIACLYTOOGKGTNODAMLVNENCSRDYVLCYFPOH
GPRGHVSKRVKRMPEFTSTOLKTTSGTROSSKNGINAPTCVDEQCEQLCEK
DEKLPPEMTPLKRALKTCQOMDKELKMPRTINCSCSTSTVVIKOGDLVNGNG
DSRAVALATRODANALVAVOLTIDLPKDPESARINHCGRPALODEPVARWMLRN
SDSPGLAMARAFEDPCLKIDYLSVDINRHLERDQYIILATDGVADYVNLKLEAD
IVASAPSDTAAVAVDOTAVRAMRLKPTNPTVPMETKEKEKPESCIEKKTTLAEC
HSHESEESTITSSKDADKEKASTNETVPMETKEKEKPESCIEKKTTLAEC
ISVKDEBMSALFGLTRVNSLSIPRFSELSNSSWRKWL"
/ complement(10192..11349)
/ gene="F309.4"
/ complement(join(10192..10323,10518..10562,10642..10742,
10834..11039,11135..11349))
/ gene="F309.4"
/ note="ESTs gb|F15498, gb|H37515, gb|T41906, gb|T22448,
gb|W43356 and gb|T20739 come from this gene."
/ evidence-not_experimental
/ codon_start=1
/ protein_id="A034675.1"
/ db_xref="GI:4966344"
/ translation="MASSSDSMWRAYNEAKLSEINGATISESSSNVTPDQORAS
AIRKLTIRGNKIDSLQSLAETHGKPRISKEKMRNRDYGRLRSKANOANALNMSN
FARSDILGPDIPDMSRVGTMDNGIVGQROVMREDEGLEDELEGTVMSTKHLA
LVAEEIDLOTRLIDLDYHVDYDSRLRRVQSLAVNMNMRSGSCMSMLLSVLGI
VGLAVYIWMLVKYM"
11967..13706
/ gene="F309.5"
/ join(11967..12044,12084..12224,12408..12954,13133..13706)
/ gene="F309.5"
/ note="Contains two Pf101344 Kelch motif domains."
/ evidence-not_experimental
/ codon_start=1
/ protein_id="A034677.1"
/ db_xref="GI:4966346"
/ translation="MPVSSVSLPCQNPFLSDFSVSIIVNGVSGKVINFEARFNC
SKASRRGSEDTPTISRVRGSLIPDLALRCLAKLSHGHYGLVECSRGMRDLVRA
DYSCYKARNGMSMLFLTERSKNOMVADPEADRMHPLPRTAVODGNHSGFACV
CVSNCLLVIGGCYAPSVSSPPIQKPYTKDMRDEPKQMKMVAISMRRPRTFACS
VSGKVVAGGRNLTHSRGIDSAEVDYDVAWRMELRAMRPOKDCSLSTRGFHYS
DOYGRABONSSEVFNPRDNTWSTVEDVWPPSRAMQRAVOVMKNDRYTTIVMGSBLIK
TRTDGEWYNGVSVSVLPNHPRELEAGYFPAALRNELVIGCVLWMBESGAGR
FDIVRLPVVACNPLDRPLMWRPKPCIPAGSIIIGCVLSLESSSP"
/ complement(14717..17027)
/ gene="F309.6"
/ complement(join(14717..15924,15993..16097,16178..17027))
/ gene="F309.6"
/ note="Similar to gb|A012423 wall-associated kinase 2
from Arabidopsis thaliana."
/ codon_start=1
/ evidence-not_experimental
/ protein_id="A034678.1"
/ db_xref="GI:4966347"
/ translation="MGVDARFLVLMILLRICGYAASFPLALRNCSDHGVNVPY
PFGIGCYKKNWFEIYCKSSSDQDPLILPRLRAVTSFLNLDPSISVYNKFYIOS
PLKHSCPNBDGYSSSLNKGSPFTSENNKTAVGCNKAAMNATGVLOICETFC
GNEIRSYKGNANTSCVYGKCCOMTIPPLLOVADVATEKREKOGVOVFLFOFITS
GSLFTPELMEYSEYTTIELMRLLDSYMSKRVLCNKNRFFEDSCSHNCEBGP
YIFGCGODIDECRDPHILNKCGKRCYANVUSICEKTPWPIILSGTSLSLILIFGM
LICKANRRKRVAAOKRRFQORNGLLQOOTSLFHSVNTKVFSSNDLENADIVFWM
SRILGOGGQGVYKMLDGMIVAVKSKALKLEENEEFINELIILSQIHNRVVKIL
GCLLETPILVEYFIPNRILFDHLNPSDEPFMSWEVEREICEVADVLSHSAVS
IPYHNDVSTNLTDEKHKRAKVSDFGISVAIDTHTLTITVOGTIGYVDPELQSN

```

HFTGKSDVYSEGVILLTELLTGKRPVSLTROEYRMIGAYFLPMARNDRRLHEIDARX  
 EECDEREYLVAKIARCLSLNSHNRTPHNDYIELDRPMKSKRGTOQAQNEENH  
 IQIAMPSSSLSSPNIVENSSFLDTRPLMPHKQ"

```

gene
18899..23154
/ gene="F309.7"
/ join(18899..21385,21458..21538,21719..21871,21997..22160,
22271..22418,22491..22613,22684..22771,22955..23154)
/ gene="F309.7"
/ note="Contains Pf100069 Eukaryotic protein kinase domain.
ESTs gb|H37741, gb|T43005 and gb|A1100340 come from this
gene."
/ codon_start=1
/ evidence-not_experimental
/ protein_id="A034679.1"
/ db_xref="GI:4966348"
/ translation="MDNRPRPFPPOHAMEPGVYVNDVPQGFPPDQGLSNANVRP
ADVPCGHYSIOTGEESFLFDRVTSQSNAPVIAQDINTPGVYHAGSEFSGV
SRMSWNGIRFQERTRTPVHEGKNGKIGHSAPEASFCODRSLGNHGVASSASG
LTKAVYLCSEFGKILPRPDLSKLRYVGETHIIISTKDSMDELKRYLEYIRTHY
VKTQFGEIDALVSCDEDLNMMEYEMENRKGOSQKLRMLFVSVDLALLGV
NKSDVSEFQYVAVNDMDLGSNSNTLNGLDSSANNLAEIDVRNTEGINVGPOL
TGDIQPSQMSQESAPPTSAQYPOSIPIHNGAFQFOQAVPATILOYAPNSPSSV
HYPOSILPNSTLOYPOSISSSYGLYPQYXGTEOFPQOYHDNNSNYSIPPEPGP
YHPHGTQONAPVPOVEEPMKPEKTVKVDYEPREPHITATNHNPNQADTETKNEP
SVATYPSQAAHMLPRDRONTPKPSTTYNDAYTBQVPSGEDQSTSGTCG
LVHDSNSLIDLDYPPDLPTRVTRSEIIPBQMLEMLRLKSDLSQSLMSP
QASTQOQEPKAEAGVISHEDSHLVNDENISGVVANSNEFLDKRTVSGGIEEARL
SHVTERSHDIPKQTSGLVIDINDREFDDPSEIIFKALSDMSGAPYOHDAQV
SVLENVHNDKNMSYFRLNADQESDQVYIIRTPQFPMDEMGELIARLHOVAPLT
INENDPQMKYTESDEEPMAVENLRTSCBEDERSKERNAGLPVPGSLADYDTSGIO
IINDDQLEKELGSGEGRVYHGKMGSPVALIKRIKSGFAGRSQDELTGEFNGE
AETLSKIHNNVYAFGVYKDGCGATLATYETELVSDSLNHLVLRKRHLDKRRRLIT
AMDAEFMEYELHAKNIVHFDKCDNLVLADSRPICKVGDGLSKTRKNTLVSGV
RGTLPMWAPRLNGSSSKSEKVDYSEFGLVLEILGEPVANMHGALIGGIYNT
LRPTISYSCSDMRILMEECWAPNPTAPSTFIAGRLRVSTAAISNOSKPPAHKAS
K"
/ complement(23491..25496)
/ gene="F309.8"
/ complement(join(23491..23643,23824..23922,24023..24104,
24189..24392,24473..24795,24882..25496))
/ gene="F309.8"
/ note="Similar to gb|L13612 DEAD-box protein (dbp45A) from
Drosophila melanogaster and is a member of Pf100270
DEAD/DBAH box helicase family."
/ codon_start=1
/ evidence-not_experimental
/ protein_id="A034681.1"
/ db_xref="GI:4966350"
/ translation="MEEPPEEGGGITINSSKRNKPTVYNVIOSTOKIDNPNPQEK
FTNPNPSDITSTINFEGLSLAENAVETCELMKRPPTVQTCYKRIILAGRVGLA
OTSGKTAARALPILRLADPYPVLELVTPRELAFQALQFKAQSLANDKCSVT
VGGMDMLTQMSLVSRPHVITTPGRKIVLENNPVPVPSFKELVLEADARVLDY
GPODELTTFIOCLPKSKOTLIFSATMSINQALLESNKAAYEAYEGLKTYDTLQ
OFIPEEDKARELYVHLSQEMEDGISANITPSTGTCORLSIMLEDELEVENIAMS
LNSQSMRLSLSKFSKGVPIILATDVASGLDIPYDVLINVDIRDRPDYVHRGR

```

alignment\_scores:
 Quality: 15.00 Length: 15
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

Alignment\_block:
 US-09-805-550-4 x F309 ..

349 CysAparGAGnglucLleuLaLaanYrLleuLgLnHs 363
 |||||||
 836 TGTGACCGTAACGAGGAATGTGCTCAACTATCTAGACAC 880
 seq\_name: gb\_pl:LES243875
 seq\_documentation\_block:

LOCUS LES243875 1498 bp mRNA linear PLN 28-JUL-1999  
 DEFINITION Lycopersicon esculentum mRNA for RAD23 homolog.  
 ACCESSION AJ243875  
 VERSION AJ243875.1 GI:5640110  
 KEYWORDS RAD23 gene; RAD23 protein.  
 SOURCE tomato.  
 ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.  
 REFERENCE 1 (bases 1 to 1498)  
 Lemaitre-Chamley M., Petit J., Raymond P. and Chevallier C.  
 Analysis of gene expression during early tomato fruit development by mRNA differential display  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1498)  
 Lemaitre-Chamley M.  
 Direct Submission  
 Submitted (23-JUL-1999) Lemaitre-Chamley M., Unite de Physiologie vegetale, INRA-Bordeaux, B.P. 81, 33883 Villenave d'Ornon cedex, FRANCE  
 FEATURES  
 source Location/Qualifiers  
 1. 1498  
 /organism="Lycopersicon esculentum"  
 /variety="cerasiformae"  
 /cultivar="West Virginia 106"  
 /db\_xref="taxon:4081"  
 /clone="Lerad23-8"  
 /tissue\_type="fruit"  
 /dev\_stage="6-10 DNA"  
 108..1277  
 /gene="rad23"  
 108..1277  
 /function="involved in DNA excision repair"  
 /codon\_start=1  
 /product="RAD23 protein"  
 /protein\_id="CAB51544.1"  
 /db\_xref="GI:5640111"  
 /translation="MKIFKTLKGFHEIEVKPDSVDYADVKNIYESVGODVPAAQO  
 MLTGGKVDKDTLLEKVAENSPVYIMLSKNVSSSTGSSISALSNTPQDSTO  
 QAROTTTTPQATLALPOSASASAPTPAPPAASSVDYDQASNLVAGSNLETTVO  
 QILDMGGSSWDRTVRLRALRAAYNPENAVDYLSGIEPTEIPPARAPAPATAP  
 ASAOINPAQDASQALAVPSSGNANMLDPEQCLTNAGSNAGNLDPLNNSPQFA  
 LRAVQANPQILQPLQELGKPNHMLRLEIODEHPDLRLINEPVEGCVLNGOTAGA  
 IPQAVTYTPEREERLERLEAMGPDRLVLEYFPACNKEELAAVYLDHLHEPDE"  
 BASE COUNT 416 a 334 c 340 g 408 t  
 ORIGIN  
 alignment\_scores:  
 quality: 14.00 length: 14  
 ratio: 1.000 gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-805-550-4 x LES243875 ..  
 Align seg 1/1 to: LES243875 from: 1 to: 1498  
 167 ArgAlaLeuArgAlaAlaTyrAsnAsnProGluArgAlaVal 180  
 ||||||||||||||||||||||||||||||||||||||||  
 636 CGGCGTTTCGTCGACATATACATCCAGAGAGCTGTT 677  
 seq\_name: gb\_p1:T24D18  
 seq\_documentation\_block:  
 LOCUS T24D18 80442 bp DNA linear PLN 16-DEC-1999  
 DEFINITION Arabidopsis thaliana chromosome 1 BAC T24D18 sequence, complete  
 sequence.  
 ACCESSION AC010924  
 VERSION AC010924.2 GI:6272370

KEYWORDS HTG.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 80442)  
 Liu, S.X., Yu, G., Sakano, H., Jha, V., Lee, J.M., Lenz, C., Pham, P., Tortum, M., Chin, J., Choi, E., Chung, M., Gonzalez, A., Hong, B., Koo, T., Li, J., Liu, A., Vaynsberg, M., Altafi, H., Brooks, S., Buehler, E., Chao, Q., Conn, L., Conway, A., Hansen, N., Johnson-Hopson, C., Khan, S., Kim, C., Lam, B., Nguyen, M., Palm, C., Shinn, P., Tambunga, G., Davis, R.W., Becker, J.R., Federspiel, N.A. and Theologis, A.  
 The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 80442)  
 Theologis, A.  
 Direct Submission  
 Submitted (28-SEP-1999) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA  
 3 (bases 1 to 80442)  
 Theologis, A.  
 Direct Submission  
 Submitted (06-NOV-1999) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA  
 4 (bases 1 to 80442)  
 Theologis, A.  
 Direct Submission  
 Submitted (16-DEC-1999) Plant Gene Expression Center, 800 Buchanan St., Albany, CA 94710, USA  
 On Nov 6, 1999 this sequence version replaced gi:5931388.  
 The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1. The sequence does not represent the sequence of the entire insert of this clone. It is shorter by 1381 bp because we submit only the unique sequence of the clone. However, in order to facilitate the joining of overlapping clones in the future for creation of larger contigs, we provide small overlaps (200 bp) between overlapping submitted clones. The 3' end of this sequence overlaps by 200 bp the 5' end of the sequence of the clone F309.  
 location/Qualifiers  
 1. 80442  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="1"  
 /clone="T24D18"  
 1202..3974  
 /gene="T24D18.1"  
 join(1202..2102,2184..2365,2479..2568,2682..2969,3324..3536,3744..3974)  
 /gene="T24D18.1"  
 /note="contains similarity to gb|AF136530 transcriptional regulator from Zea mays. ESTs gb|F14071, gb|Z28623, gb|A198935 come from this gene."  
 /codon\_start=1  
 /evidence-not\_experimental  
 /protein\_id="AAFI8488.1"  
 /db\_xref="GI:6587797"  
 /translation="WSISDEAEISESIEDYSETPYLLRDGYTKVKNQRCPC  
 AGKKQDYKRELYAHATGVSGSATRSALQKANTLAMPLENLALAYAEVPPPV  
 VPPQDETEPNHNHYVPMGIIVNPELEADKDLILDSAWIQTLEKFRPEYNAF  
 VPEODSIVGLAKFNGDSGFAGATELEKEPTQSSKKEWTERGDSERAYGCAR  
 ADDEFSQGIYEYSLEKGLRVSQISQKNQDRTVLEISDMTAMNEDLNKQVYS  
 VNRMTSIOBYIDEKKNLHQAAPADRTKKQOMSLPHIKLILYDKLSNEIDRRKRD  
 ESRKQLEKHEALTELDKQDEDRKSDANKSLQLASREOKKADSEVLRIYERHQR  
 OKEDALNKILLEKQDLYKQTEIMEIQELKGLQVYKHLGDDDDAVOKKKKNEDEL  
 DDKKAELEKLESMSNVLMTEKROSNDEIOAKRKLIAGITGLGAEITDIGYRMGELD  
 EKPEFDVCKLRYSANAEVAATLTCSTQENLKNKSMQPFKEGTGDAEYVDDDE  
 DKTKLREMGKREVNHAVALVENMENYASGTYTTPPELNFREGRKATLKEVITFISN  
 DIKILKRRRT"  
 5213..6073  
 /gene="T24D18.2"

gene

```

CDS
5213..6073
/feature="T24D18.2"
/note="Similar to g11060809 CCR4-associated factor 1
(gCAF1) from Mus musculus. EST gb|226822 comes from this
gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AA18489.1"
/db_xref="GI:6587798"
/translation="MSQAPNPDEEDDTIEIREYWNHNLDEOMALLIEOSIDPFYVAD
TEPPGIVCKTVANPNPNPYSIHREYNIDLKANVNLKIQGLTISDEKMLPTGC
TNKOCIMOFNREFREYNSIDFALDSIELKSAIDLEKNEGCDARFPEMLGSGSV
VLNDKIHMTWHEHCGDYGVLKILLSGKEIPELISDPDOMEREPYVDIKYLMGRT
NIXGLKELIAELGVKRVGISHQAGSLSLTLTFTKMEFFPTGSLIKSGFLGID
NPRLLTGSKN"
/feature="T24D18.3"
/join(6953..6960,7349..7649,7771..7836,7933..7992)
/feature="T24D18.3"
/note="Strong similarity to gblAF067732 ribosomal protein
S12 from Hordeum vulgare. ESTs gb|741772, gb|742570,
gb|A1999345, gb|720784, gb|F20068 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AA18490.1"
/db_xref="GI:6587799"
/translation="MSGDEAPVVPVPPVPAEPAIDEDMLTALLETTLKARAYGV
VRLGHECAKLIKERVAVLVLAEDCNQDPVVKLVKACDHEVRLLVPSAKTLGEWA
GLCKIDSEGNARVAGVCSCLVVKDFGEETLALIVNKHIAQ"
/feature="T24D18.4"
/join(8923..9036,9346..9549,9665..9846,9939..10065,
10150..10347,10469..10519,10604..11311,11396..11493,
11579..11652,11736..11787,11877..11963,12052..12095,
12184..12258,12341..13301)
/feature="T24D18.4"
/codon_start=1
/evidence=not_experimental
/product="T24D18.4"
/protein_id="AA18491.1"
/db_xref="GI:6587800"
/translation="MGLPVEATFEKSKQALTDAAENLAKPHSTDAITLLEWESIL
AYEODLSVVOKALHPMRAIVSADLRMPDSVAVSVSCLEIMRTLAAPAPASD
EOMKDIPOVTEFAPEKLIADASSRYKAEYILETVAKVRSILMLDECDLVLEMO
RFLKTRIPDHPOLVLYMETIMITVIDESEVPMDELITLVTKKDSOVSAATL
VEVYLSSCTCKLQPCIMEALSSGTSIDMTSPVSSICSEFAITQAHNDKCRDNEA
DEKISGOVVPVSDLEDKLNGLSRKGTSGRSGARAGTTRANDDEVITANELEST
DAETASGSTRKRGKMPKSLMNPREGYSFRTSSSKVOEKELGDSISGVAKKVPVPS
KVGOTNOSVVISLSSSGARATSGRSRTRKMEETDDVSVATOPAKKOTVKTNPAAK
EDLTRNKKHEDGKITCKSSKREKADNGLAKTSAKKPLAETMMVKRSGKLVHSDAK
KNSGASMDPTLPQSSSKKKKDSRATTPATKSEQAPKSHPMKRIAGEVEVSNTE
LGEELVGRVNWVWMLDKKFEYGVIKSYCVKTMHQVTVSDGVEELNKEFEKIE
DKSSASEDEDLESTPLSAFIQREKSKRKIVSNVESSSPEYVSSMOKKDKDS
VDSIKOTRKTKALKAVSNPESTGKNLNSKILNGEEDTKRGRTGKOKXOYAOAH
RKIEKDCDEOEDLEKDEEDSLKIKESAPEDRMDOHLPENHNVETRTDDEBEA
AKEPPTAESTNGCEPNAEPETDCKEHSKLEPNAEKRSODEEBAKPEPAELKTGCE
NQBAKELTAERKTDEEHRVADVEOKSDKETNVEPEGEOKKVEEENAPKPKYEL
EKESEAKEQYADTKLIEKEDMSKTEEDIKETYSIPETGKGVNEAEDDQVIREL
EESSDRAEVSTVLEVP"
/feature="T24D18.5"
/join(14106..14238,14360..14514,15043..15228,15405..15757,
16959..17166)
/feature="T24D18.5"
/note="Strong similarity to cinamoyl CoA reductase
g112960364 from Populus balsamifera. ESTs gb|N95902,
gb|A1992693, gb|A1995837 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AA18492.1"
/db_xref="GI:6587801"
/translation="MPVDVASPAKGTVCVAGGYIASWIVKILLEGYTVKGTVRNP
DPRKNHMLRELBGGRKRLILCKADLDQYELAKAIDGCGVPHITAPRVYDDPQMWEP

```

```

gene
complement(17453..20317)
/feature="T24D18.6"
/join(17453..17770,17863..17956,18041..18172,
18248..18352,18449..18622,18698..18767,18839..19020,
19098..19207,19342..19431,19565..19643,19731..19771,
19858..19936,20225..20317)
/feature="T24D18.6"
/note="Strong similarity to g112130080 Nramp1 protein from
Oryza sativa and contains a Pf01566 Nramp domain."
/codon_start=1
/evidence=not_experimental
/protein_id="AA18493.1"
/db_xref="GI:6587802"
/translation="MAETRASGNSRISNPILNDSNOILVPEKSKMKNFYSYLP
GFLVSLAYIDPGEFEDLQSGAQRYKELMILIVASCAALVQSLANLGVGKHLA
EHCRAEYSKVPNFMVWVAVIALVAFALMNFENPVMVGLTGLSTILIAL
QOYGRKLEFLIAFLVETLALCFEVELHYKPPDKELYGLFVPOLGNGATLAIL
LGAVMVPHNLELSALVLSKRIPRSYGIKACRYLIEGSLAMVAFLVNSVYS
GAYCNSDLSPEPRAQODDLNKASFLRNVYKSKSRFPAIAILASGSSITTCY
AGQYVMOGFLDLRLERPLRNFRLCLAITPSLIVALIGSAGAKLITIASMTLSEL
PFALVPLKFTSSKTKGSHANSVLVSSVWIIIGGLMGINITYLVSSFKILLHSHM
NLVAIVFLGVLAGFSGIATVLAALSYLVLRNKRNSSTHFLDFNSQTEETLPREDIAN
MOLPNRAVIGDNL"
/feature="T24D18.7"
/complement(21795..23745)
/feature="T24D18.7"
/join(21795..22106,22210..22351,22466..22549,
22637..22694,23283..23745)
/feature="T24D18.7"
/note="Contains similarity to g1112785
DNA-3-methyladenine glycosylase I from Escherichia coli."
/codon_start=1
/evidence=not_experimental
/protein_id="AA18494.1"
/db_xref="GI:6587803"
/translation="MSVPPFRSIVNSDEREFRSVLGPTGNKLRKPPGMLEKPMMEK
TIIIDSKDEKAKKPTTPASPRTLTKOCSLSILIRKNSAMTASYSDDSSCESSPL
SVASSSSCKKVVVRSGSVSTRLSLQKKEEKYSGDGFADGRCAWITPKADPCYVA
FHEBNGVYPHNDKRLPELCLISGALAEISWTIDLSRHLRLKRVFDPFVAAEIND
KLTLPGETAAILSLSEVKIRSIIDNSHVKIILAECSLKKYMNMFVNKPIYOSQRY
QROVPVTKTSAEELISDLVRRGRSVSPVIVSFMQAGLTNDHLIGCFRYODCCYDA
ETTTTTRAKKKNRESDK"
/feature="T24D18.8"
/join(24565..24989,25088..25238,25326..25516,25589..25846,
25946..26196)

```

## alignment\_scores:

Quality:	14.00	Length:	14
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

## alignment\_block:

US-09-805-550-4 x T24D18 ..

Align seg 1/1 to: T24D18 from: 1 to: 80442

```

167 ARGAlaLeuATGAlaAlaTyrTAsnAsnProGluATGAlaAlaVal 180
|||||
79856 CGTGCACCTCGTGTGAGCATATPAAACCTGTGAGAGCAGTG 79897
|||||

```

---

